Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

aS494
.5
.E8A8

TECHNICAL MEMORANDUM NO. 12

THE EFFECT SYSTEMATIC STRATIFIED SAMPLING HAS ON THE COMPUTATION OF VEGETATIVE INDEX NUMBERS

FOR CROP CONDITION ASSESSMENT DIVISION

UNITED STATES DEPARTMENT OF AGRICULTURE
FOREIGN AGRICULTURAL SERVICE

HOUSTON, TEXAS

UNITED STATES DEPARTMENT OF AGRICULTURE
FOREIGN AGRICULTURE SERVICE

THE EFFECT SYSTEMATIC STRATIFIED SAMPLING HAS ON THE COMPUTATION OF VEGETATIVE INDEX NUMBERS	FIRST ISSUE

APPROVED BY:

James R. Hickman

James R. Hickman, Crop Condition Assessment Division

1. REASON FOR ISSUANCE

Documents the effects a systematic stratified sampling design of the Landsat data has on the computation of the Ashburn Vegetative Index (AVI) and the percentage of positive AVI picture elements (PPAPE).

2. COVERAGE

The paper evaluates the effects a 3 scan line by 3 pixel and a 5 scan line by 5 pixel sampling of the Landsat digital data have on the computation of the AVI and the percentage of positive AVI picture elements. The study area was in northeastern Montana.

3. ACKNOWLEDGEMENT

Mr. Elmo Dario for his computer programming support and Mr. Pat Ashburn for his support in sampling design and results evaluation. Their contributions are hereby gratefully acknowledged.

4. PREPARED BY: Andrew C Aaronson DATE: Sept 3 1980
Andrew C. Aaronson - FAS, CCAD

TABLE OF CONTENTS

INTRODUCTION

1.1 SUMMARY AND CONCLUSIONS

Due to the large volume of Landsat data which is routinely processed by the Crop Condition Assessment Division (CCAD) (850 full frames during peak processing) a sampling method was required to reduce the data load and improve computer efficiency without sacrificing the loss of important crop condition related information. The CCAD chose to impose a 5 scan line by 5 pixel sampling of the Landsat data for the computation of vegetative index number components used in crop condition analysis. The decision to implement this logic was based on a number of factors including computer efficiency, data storage and flow. Additionally, a 5 scan line by 5 pixel sampled Landsat image is compatible with other CCAD image processing and display systems.

The purpose of the study was to assess the accuracy of the estimates of the average AVI and the percentage of positive AVI picture elements (PPAPE) for variable sized areas from the implementation of a number of scan line and pixel sampling factors.

Correlation analysis was selected for evaluating the accuracy of the estimates of the average AVI and PPAPE obtained from sampled Landsat data for each of three sized images (5 by 5, 10 by 10 and 15 by 15 mile areas) located in northeastern Montana. A 3 scan line by 3 pixel and a 5 scan line by 5 pixel sampling design were selected for testing.

The results indicated that a 3 by 3 or 5 by 5 scan line and pixel sampling resulted in estimates of the average AVI and the PPAPE that were not significantly different from those computed using full resolution data for each of the three sized areas in the test region.

1.2 BACKGROUND

The Crop Condition Assessment Division (CCAD) of the Foreign Agricultural Service (FAS) is responsible for verifying and assessing the impact of adverse conditions on crops in important foreign producing areas and reporting the results to FAS commodity analysts in Washington, D.C. The CCAD uses Landsat satellite data as a primary information source for crop condition assessment. Figure 1 shows the distribution of candidate foreign agricultural producing areas presently assessed by the CCAD. These areas require coverage from 450 to 850 Landsat frames depending on the time of year. The variability in Landsat frame coverage during the year is a function of latitude and crop growing seasons. During the May through August time period, the CCAD experiences a

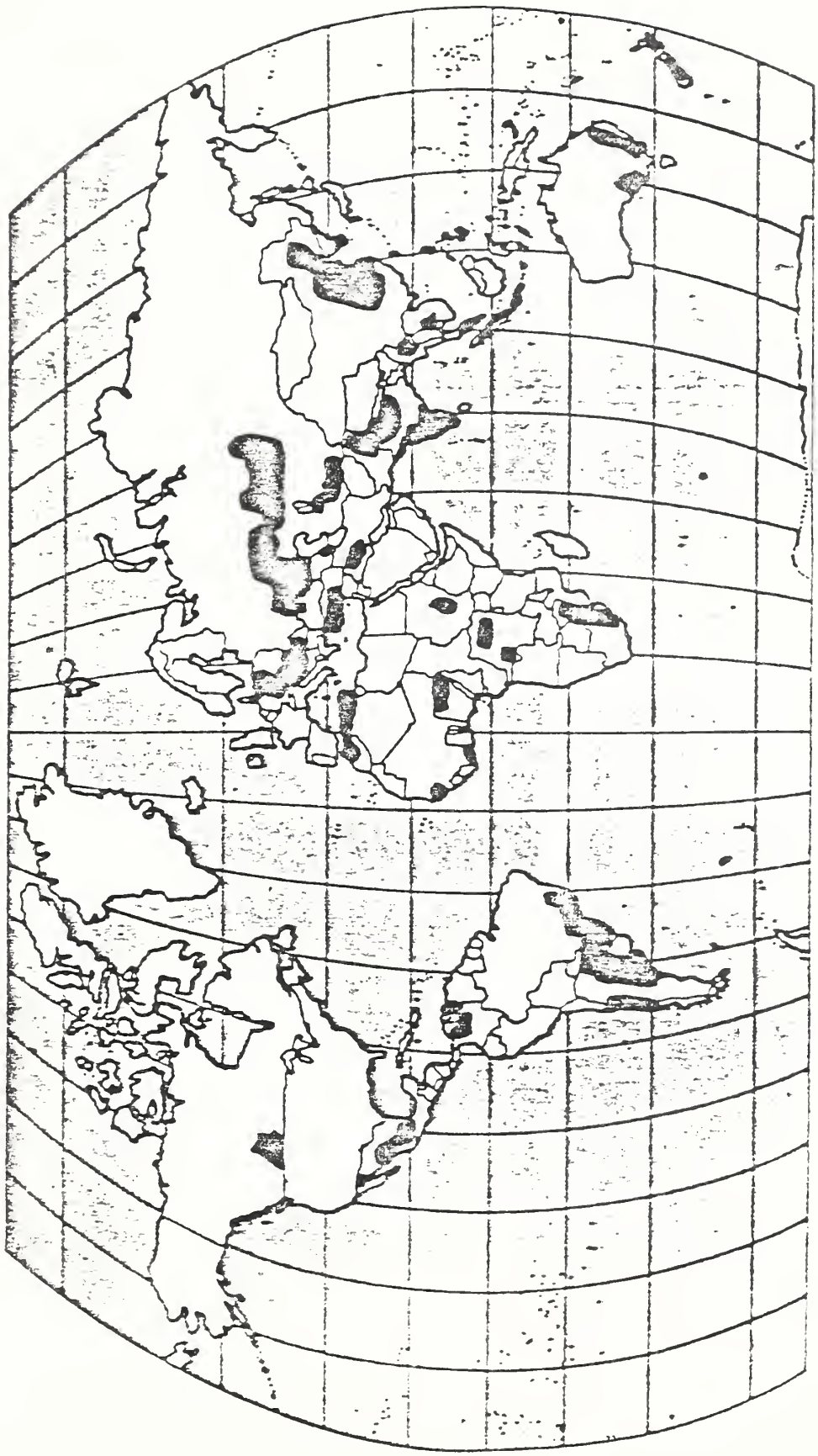

FIGURE 1-1 CCAD IS RESPONSIBLE FOR ASSESSING MANY FOREIGN AGRICULTURAL PRODUCING AREAS

peak in Landsat frame coverage because it is acquiring data from both northern and southern hemisphere countries.

Methods used by the CCAD for assessing crop conditions using Landsat multispectral scanner (MSS) data include visual screening of the imagery and computer analysis of the digital data. The CCAD uses MSS digital data in determining the areal extent of an abnormal event affecting agricultural production and in assessing crop conditions. Vegetative Index Numbers (VINs) are computed from the MSS digital data and are used by CCAD analysts to measure relative "greenness" of agricultural areas. Researchers have developed a number of VINs all of which are sensitive to the density of the vegetative ground cover. The basis for development of the VINs was to reduce the four Landsat spectral bands to their greenness component. The greater the density or canopy of vegetative areas the greater the VIN. CCAD analysts make qualitative measurements of crop condition by comparing VINs computed from different areas of a region or country and/or from the same area for different years. It has been shown[1,2] that a high correlation exists among the available VINs. The CCAD chose to process the Ashburn Vegetative Index* (AVI) based primarily on its computation efficiency.

The computation of the AVI was integrated[3] with the CCAD world-wide gridded data reference system. This system defines unique geographic 25 by 25 mile grid cells in which a variety of data sources are stored. The average AVI and the percentage of positive AVI picture elements (PPAPE) routinely are computed for each grid cell entirely contained in a Landsat full frame acquisition and stored in the geographic data reference system. This data is then available to the analyst for crop condition analysis. In total the CCAD geographic data reference system provides for the computation and stroage of the average AVI and the PPAPE for over 180,000 grid cells distributed around the world.

Due to the large volume of Landsat data which is routinely processed by the CCAD a sampling method was required to reduce the data load and improve computer efficiency without sacrificing the loss of important crop condition related information. The CCAD reviewed many different sampling approaches. A sample based on agricultural production probabilities was not possible due to a severe limitation of the required historical production statistics to compute the sample. It was felt a systematic stratified sample would provide the best results and provide the CCAD an unbiased estimate of the AVI and the PPAPE for each grid cell. A number of

*AVI=BAND 7 - BAND 5

systematic scan line and pixel sampling factors were reviewed as a method for significantly reducing the data load. Figure 2 illustrates the results of imposing a systematic stratified 2 scan line by 2 pixel sampling on a data set. The resulting sample product represents approximately 23 percent of the original data set.

The CCAD chose to impose a 5 scan line by 5 pixel sampling of the Landsat data for the computations of the VINs. The decision to implement this logic was based on a number of factors including computer efficiency, data storage and flow. Additionally, a 5 scan line by 5 pixel sampled Landsat image is compatible with other CCAD image processing and display systems.

1.3 PURPOSE

The purpose of this study was to assess the accuracy of the estimates of the average AVI and the PPAPE for variable sized areas from the implementation of a number of scan line and pixel sample designs.

1.4 DATA SET

Landsat full resolution images 5 by 5, 10 by 10 and 15 by 15 miles in size were extracted from a quarter section of a full frame located in northeastern Montana. The final data set included 70 - 5 by 5 mile areas; 27 - 10 by 10 mile areas; and 12 - 15 by 15 mile areas.

1.5 APPROACH

Correlation analysis was selected for assessing the accuracy of the estimates of the average AVI and PPAPE obtained from the sampled Landsat data for each of the three sized images. The values estimated from the sampled Landsat data were correlated to those values generated from the full resolution data. The average AVI and PPAPE were computed for each 5 by 5, 10 by 10 and 15 by 15 mile area contained in the quarter section Landsat full frame using full resolution data, a 3 by 3, and a 5 by 5 scan line and pixel sampling.

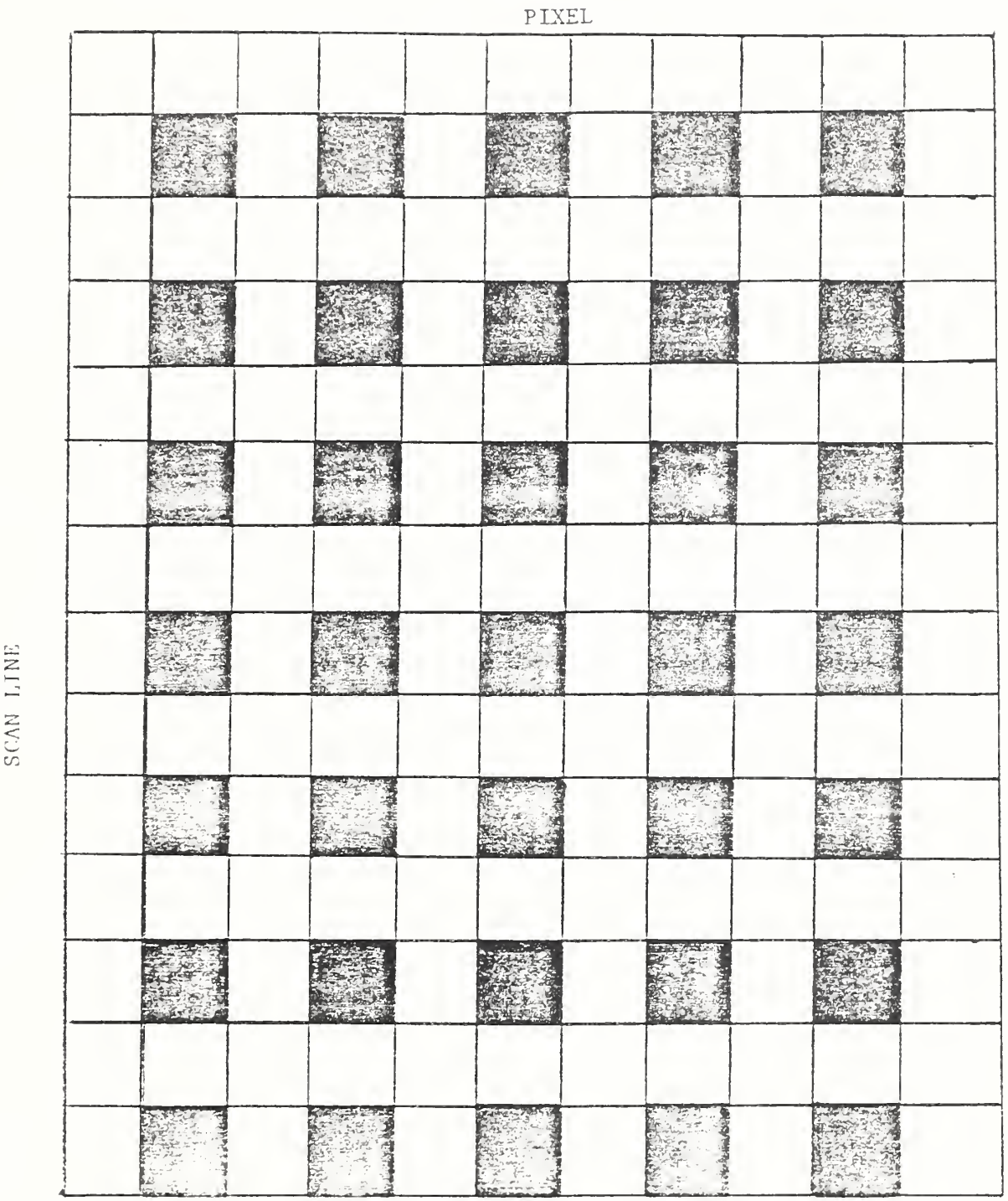

FIGURE 1-2 SAMPLED PICTURE ELEMENTS FROM A 2 SCAN LINE BY 2 PIXEL SAMPLING FACTOR

2.0 CORRELATION ANLAYSIS RESULTS

Tables 2.1 through 2.6 summarize the results of the correlations of the average AVI and the PPAPE computed between the full resolution data and the sampled data for the 5 by 5, 10 by 10 and 15 by 15 mile areas. The correlation coefficients computed were .97 and greater for the average AVI and the percentage of positive AVI picture elements for each of the three sized areas.

The imposition of a 3 by 3 or 5 by 5 scan line and pixel sampling resulted in values that were not significantly different from those computed using the full resolution data. This systematic stratified sampling approach used in this data set is unbiased and accurately represents the vegetation contained in a given area. The high correlations obtained even with a 5 scan line by 5 pixel sampling are due to the inclusion of enough representative picture elements of each land unit contained in a given area. For example, a land unit may be a forested, rangeland, or cropland area. The more homogeneous an area in terms of its land units, the greater the probability of a high correlation between the values computed from full resolution data and those computed from sampled data. Conversely, the smaller the land units and the more varied the conditions within a land area the more likely a decrease in the correlations. This can only be concluded after an in-depth study of the impact varying sampling factors have on the values using Landsat data acquired over variable conditions found in foreign areas. Certainly the results obtained in this study support the hypothesis that the correlations will be high even in the more heterogeneous areas.

TABLE 2.1 CORRELATION COEFFICIENTS FOR THE AVERAGE AVI, 5 BY 5 MILE AREAS

SKIP FACTOR	0,0	3,3	5,5
0,0	1.00	.98	.99
3,3	.98	1.00	.97
5,5	.99	.97	1.00

TABLE 2.2 CORRELATION COEFFICIENTS FOR THE PPAPE, 5 BY 5 MILE AREAS

SKIP FACTOR	0,0	3,3	5,5
0,0	1.00	.99	.99
3,3	.99	1.00	.98
5,5	.99	.98	1.00

TABLE 2.3 CORRELATION COEFFICIENTS FOR THE AVERAGE AVI, 10 BY 10 MILE AREAS

SKIP FACTOR	0,0	3,3	5,5
0,0	1.00	.99	.99
3,3	.99	1.00	.98
5,5	.99	.98	1.00

TABLE 2.4 CORRELATION COEFFICIENTS FOR THE PPAPE, 10 BY 10 MILE AREAS

SKIP FACTOR	0,0	3,3	5,5
0,0	1.00	1.00	1.00
3,3	1.00	1.00	.99
5,5	1.00	.99	1.00

TABLE 2.5 CORRELATION COEFFICIENTS FOR THE AVERAGE AVI, 15 BY 15 MILE AREAS

SKIP FACTOR	0,0	3,3	5,5
0,0	1.00	.99	1.00
3,3	.99	1.00	.98
5,5	1.00	.98	1.00

TABLE 2.6 CORRELATION COEFFICIENTS FOR THE PPAPE, 15 BY 15 MILE AREAS

SKIP FACTOR	0,0	3,3	5,5
0,0	1.00	1.00	1.00
3,3	1.00	1.00	.99
5,5	1.00	.99	1.00

NATIONAL AGRICULTURAL LIBRARY
1022467914

EXHIBIT 1
(PAR. 1.2)

REFERENCES

A. Aaronson, A. C., Davis, L. L., and May, G. A., "Results of the Vegetative Index Correlation Study", Crop Condition Assessment Division, Foreign Agricultural Service (6-TM, Feb. 9, 1979).

B. Aaronson, A. C., Davis, L. L., "An Evaluation of Relationships between Vegetative Indices, Soil Moisture and Wheat Yields", Crop Condition Assessment Division, Foreign Agricultural Service (9-TM, Sept. 18, 1979).

C. Driggers, W. G., Downs, J. M., Hickman, J. R., and Packard, R. L., "Data Base Design for a Worldwide Multicrop Information System", USDA Proceedings of Technical Sessions, LACIE Symposium, Vol. II, October, 1978.

NATIONAL AGRICULTURAL LIBRARY
1022467914